Received: 13 March 2020 | Accepted: 26 June 2020

DOI: 10.1111/jvim.15851

STANDARD ARTICLE

Journal of Veterinary Internal Medicine Open Access

ACVIM American College of Veterinary Internal Medicine

Development and internal validation of a prediction tool to aid the diagnosis of Cushing's syndrome in dogs attending primary-care practice

Imogen Schofield[1] | David C. Brodbelt[1] | Stijn J. M. Niessen[2,3] | David B. Church[2] | Rebecca F. Geddes[2] | Noel Kennedy[1] | Dan G. O'Neill[1]

[1]Pathobiology and Population Sciences, The Royal Veterinary College, Hatfield, UK

[2]Clinical Science and Services, The Royal Veterinary College, Hatfield, UK

[3]The VetCT Telemedicine Hospital, St John's Innovation Centre, Cambridge, UK

Correspondence
Imogen Schofield, Pathobiology and Population Sciences, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Herts AL9 7TA, UK.
Email: ischofield6@rvc.ac.uk

Funding information
Dechra Veterinary Products Ltd

Abstract

Background: Novel methods to aid identification of dogs with spontaneous Cushing's syndrome are warranted to optimize case selection for diagnostics, avoid unnecessary testing, and ultimately aid decision-making for veterinarians.

Hypothesis/Objectives: To develop and internally validate a prediction tool for dogs receiving a diagnosis of Cushing's syndrome using primary-care electronic health records.

Animals: Three hundred and ninety-eight dogs diagnosed with Cushing's syndrome and 541 noncase dogs, tested for but not diagnosed with Cushing's syndrome, from a cohort of 905 544 dogs attending VetCompass participating practices.

Methods: A cross-sectional study design was performed. A prediction model was developed using multivariable binary logistic regression taking the demography, presenting clinical signs and some routine laboratory results into consideration. Predictive performance of each model was assessed and internally validated through bootstrap resampling. A novel clinical prediction tool was developed from the final model.

Results: The final model included predictor variables sex, age, breed, polydipsia, vomiting, potbelly/hepatomegaly, alopecia, pruritus, alkaline phosphatase, and urine specific gravity. The model demonstrated good discrimination (area under the receiver operating curve [AUROC] = 0.78 [95% CI = 0.75-0.81]; optimism-adjusted AUROC = 0.76) and calibration (C-slope = 0.86). A tool was developed from the model which calculates the predicted likelihood of a dog having Cushing's syndrome from 0% (score = −13) to 96% (score = 10).

Conclusions and Clinical Importance: A tool to predict a diagnosis of Cushing's syndrome at the point of first suspicion in dogs was developed, with good predictive

Abbreviations: ALKP, alkaline phosphatase; ALT, aminotransferase; AUROC, area under the receiver operating curve; CITL, calibration-in-the-large; EHRs, electronic health records; EPV, events-per-variable; IQR, interquartile range; LDDST, low dose dexamethasone suppression test; LRT, likelihood ratio test; UCCR, urine cortisol-creatinine ratio; USG, urine specific gravity; WHWT, West Highland white terrier.

performance. This tool can be used in practice to support decision-making and increase confidence in diagnosis.

KEYWORDS
canine, diagnosis, electronic patient record, endocrinology, hyperadrenocorticism, VetCompass

1 | INTRODUCTION

Spontaneous Cushing's syndrome (or hyperadrenocorticism) is one of the more common endocrine diseases in dogs with an estimated UK prevalence of 0.28%.[1,2] Cases of Cushing's syndrome typically show varying combinations of polydipsia, polyuria, polyphagia, muscle atrophy, hepatomegaly, dermatological changes, and laboratory changes.[1,3-5] Specific diagnostic tests such as the adrenocorticotropic hormone (ACTH) stimulation test and the low dose dexamethasone suppression test (LDDST) are commonly used to increase confidence in the diagnosis of Cushing's syndrome.[1,6] However, there is no single highly accurate test, making a confident diagnosis difficult.[7-11] The ACTH stimulation test and LDDST have low positive predictive values when used in a low prevalence setting, therefore their interpretation are reliant on a high prior index of suspicion of disease and are impractical for disease screening.[6,9] Other tests more suitable as screening tools, such as the urine cortisol-creatinine ratio (UCCR), are not commonly used in primary-care practice and are impacted by a high false positive rate with specificity estimates ranging from 21% to 77%.[1,10,12-14] Novel methods to aid the identification of the highest risk dogs within the at-risk population are warranted to increase confidence in diagnostic blood tests through an increase of the positive predictive value, avoid unnecessary testing, and to generally aid decision-making for primary-care practitioners. A timely and correct diagnosis of Cushing's syndrome is important because of the reduced quality-of-life in affected dogs and to ensure dogs are appropriately managed while living with the disease.[15]

Although individual risk factors such as age, breed, and sex have been associated with the diagnosis of Cushing's syndrome,[1,2,16] the cumulative risk and predictive value from combinations of results from these risk factors for individual dogs are unknown. The previously reported explanatory regression models provide population level inferences about the strength of a risk factor association in relation to the causal hypothesis, but these are not directly applicable to the diagnosis of Cushing's syndrome in individual dogs by practitioners in practice. Prediction models aimed at the individual level are increasingly being developed and utilized in human medicine to aid decision-making in clinical settings.[17] In a diagnostic setting, prediction models combine 2 or more risk factors to estimate the probability that a certain disease is currently present (or absent) in an individual.[18] Regression and machine learning methods have been used to develop clinical prediction models in humans with sufficiently large datasets necessary to ensure accurate predictions for diseases such as cardiovascular disease, dementia, and diabetes mellitus.[19-21]

Our study aimed to develop and internally validate a model to predict dogs receiving a diagnosis of Cushing's syndrome using demographic, presenting clinical signs and routine clinicopathologic data. From the model, it was aimed to develop a corresponding tool which calculates the predicted likelihood of a specific dog having Cushing's syndrome. This tool could be readily applied by clinicians in practice to evaluate an individual dog's risk of disease before confirmatory diagnostic testing, to increase confidence in the diagnosis of Cushing's syndrome.

2 | MATERIALS AND METHODS

2.1 | Study population and predictors

Data were collected from the VetCompass programme, which collates electronic health records (EHRs) from primary-care veterinary practices in the United Kingdom.[22] Dogs in the VetCompass cohort were required to have been under veterinary care in 2016 which was defined as having (1) at least 1 EHR recorded during 2016 and/or (2) at least 1 EHR recorded both in 2015 and 2017. Search terms were applied to EHRs of these dogs to identify those where Cushing's syndrome was considered as a clinical diagnosis: "Cushing*, HAC, hyperadren*, hyperA, trilos*, Vetory*." All dogs eligible for inclusion in the analysis were reviewed through manual revision of EHRs identified by the search terms. The case definition required dogs to have (1) an initial diagnosis of Cushing's syndrome recorded within the EHR between January 1, 2016 and June 1, 2018 and (2) a record of a LDDST or ACTH stimulation test being performed within the EHR before diagnosis. Dogs were excluded as a case if (1) a subsequent revision of the diagnosis was made in the EHR, (2) a diagnosis was made before their first health record, or (3) if cases were considered iatrogenic or had glucocorticoid administration in the 30 days before first suspicion. A comparison reference population required dogs to have (1) a recorded suspicion of Cushing's syndrome within the EHR between January 1, 2016 and June 1, 2018 as identified by the search terms, (2) subsequently had Cushing's syndrome ruled out after undergoing at least a UCCR, LDDST, and/or an ACTH stimulation test, and (3) an alternative diagnosis made within the EHR. Dogs where Cushing's syndrome continued to be suspected but the disease neither confirmed nor ruled out during the time period from January 1, 2016 to June 1, 2018 were excluded from the analysis. A random selection of eligible dogs was included in analysis, based on a priori sample size calculations. Effective sample size was estimated using events-per-variable (EPV), which is the ratio of the number of predictor variables included in model development relative to the number of events (number of dogs diagnosed with Cushing's syndrome).[23] An EPV of at least 10 is recommended and frequently cited in the literature.[24,25] An a priori sample size calculation estimated that

between 260 and 520 cases were required if 26 predictor variables were included in the modeling process, to ensure a sufficient EPV between 10 and 20. Ethics approval was provided by the Royal Veterinary College Ethics and Welfare Committee (URN SR2018-1652). All analyses were carried out using Stata 15 (Stata, College Station, Texas).

Predictor variables included from routinely collected data were age, breed, bodyweight, sex, and neuter status. Breeds were categorized according to a standardized breed list adapted from the VeNom Coding Group system (Venom Coding Group 2019). Individual breeds were specified if at least 20 dogs of that breed had been included for analysis. All other purebreds were grouped into a "purebreed other" category. Dogs classified as a breed-cross (eg, poodle X) or a designer breed (eg, cockapoo) were classified into a "crossbreed" category. Sex was categorized to include neuter status: female-entire, female-neuter, male-entire, or male-neuter. Age at first suspicion (years) was calculated by using the date of birth and date of first suspicion of Cushing's syndrome. Bodyweight (kg) was the bodyweight value recorded closest to the date of first suspicion.

Additional data were extracted manually from the EHRs of the cases and noncases. Date of first suspicion was the earliest date with evidence that Cushing's syndrome was being considered as a diagnosis. Clinical signs and laboratory measurements 1 week before and 1 week after the date of first suspicion were extracted. Animals with no recorded information regarding clinical signs within this 2-week period were excluded from the analysis. Individual clinical signs as evident in the EHRs were extracted as binary variables: "yes" or "no" (either no information recorded or specifically recorded as not present). Clinicopathologic data extracted included categorical variables of alkaline phosphatase (ALKP) and alanine aminotransferase (ALT) (recorded as "elevated," "not elevated," or "unknown"). Proteinuria (based on a urine dipstick, including a trace recording or a urine protein-creatinine ratio) was recorded as "present," "not present," or "not recorded." Urine specific gravity (USG) was recorded as "dilute" (≤1.020), "not dilute" (>1.020), or "not recorded." Continuous data for recorded ALKP enzyme activities and USG measurements were also extracted. Treatment data were extracted for insulin, l-thyroxine supplementation, and antihypertensives (amlodipine, benazepril, enalapril, or telmisartan).[26] Additional clinical management data were extracted to identify dogs that were hospitalized and had surgery for a cruciate rupture in the previous 12 months before first suspicion.[6,27] For noncases, the final alternative diagnosis recorded in the EHR was also extracted.

Data were examined before modeling to report descriptive statistics for the predictor variables. Categorical data were presented showing the counts and corresponding percentages. Quantitative data were assessed graphically for normality; normally distributed data were summarized using the mean (SD) and non-normally distributed data using the median (interquartile range [IQR] and range). Potential pairwise correlations between predictor variables were explored to identify potential collinearity using correlation coefficients for continuous predictors. Predictor variables were considered highly correlated if $r > 0.80$.[28] Associations between categorical variables were assessed by chi-squared tests and were considered to be highly related if $P < .001$ and were plausibly associated with each other.[29] When pairs of highly correlated predictor variables were identified, the variable considered to be most complete within the data set and most clinically relevant was selected for modeling.[28] Variables with large amounts of missing data (>65%) were excluded from further analysis, based on the consensus of the authors.[30,31] A separate "not recorded" category was used to include missing data for variables with ≤65% missingness.

2.2 | Model development and internal validation

Multivariable binary logistic regression with 200 bootstrap samples was used to develop and internally validate the diagnostic prediction model for Cushing's syndrome in dogs.[32,33] In each bootstrap sample, dogs were randomly selected with replacement until a data set of the same size was obtained, including approximately 63.2% of the dogs from the study population.[32] A backward stepwise model building approach was used with sequential elimination of predictors with the largest P value based on likelihood ratio tests (LRTs), within each bootstrap sample (LRT $P < .10$).[34,35] No univariable screening was undertaken. Predictors which remained significant at the 10% level were retained in the final model, to minimize the risk of rejecting predictor variables potentially important in future applications of the tool.[35]

Internal validation assessed how well the model was likely to perform in an independent data set. Developed prediction models tend to overfit the data and can be overly "optimistic" of their future performance.[33] Internal validation quantified the model optimism by: (1) estimating optimism-adjusted performance measures and (2) adjusting the model for overfitting by reducing the model coefficients toward the null (shrinkage).[23,33] The average difference between the performance of the bootstrap samples (apparent performance) and the dogs not included in the bootstrap samples calculated the optimism of the model and estimated optimism-adjusted performance measures.[36] Uniform shrinkage to correct for model optimism was applied by multiplying the optimism-adjusted calibration slope with the coefficients.[21,33] The model constant was reestimated based on the adjusted coefficients to maintain overall model calibration.[33]

Continuous variables were assessed for linear associations with the outcome using the LRT for departure from trend and LRT for extra-linear effect. Nonlinear continuous predictors were modeled using linear splines.[37,38] Potential confounding was assessed by reinserting eliminated predictors into the developed model to assess the magnitude of changes in the model coefficients. A 20% change in the odds ratio when the subsequent variable was added to the model was used to identify potentially confounding variables.[28] Potential interactions between predictors were assessed using LRTs. The potential clustering effect of the clinics included within the study was assessed by including clinic ID as a random effect in a mixed effect model.

Performance of the model was assessed by examining the calibration and discrimination.[39] Calibration measures the agreement between the observed outcomes and predictions. A calibration plot compared the predictions within each bootstrap sample with the

observed outcomes. The plot compares the mean observed proportions of dogs with a diagnosis of Cushing's to the mean predicted probabilities by deciles of predictions. Perfect predictions should lie on the 45° line.[40,41] Overall model calibration was calculated from the mean calibration plot gradients (c-slope) and intercepts (calibration-in-the-large [CITL]). The c-slope was used as the shrinkage factor to gain the optimism-adjusted model coefficients. The c-slope is often lower than 1 for models developed using relatively small data sets suggesting that predictions are too extreme (ie, low predictions are too low, high predictions are too high).[39] CITL > 0 suggests that observed proportions are higher than the predicted probabilities (predictions are systematically too low) and CITL < 0 suggests that predicted proportions are higher than the observed proportions (systematically too high).[39]

Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC), with 95% confidence intervals. The Brier score and Cragg-Uhler's (Nagelkerke) R^2 assessed overall model performance, a concept related to goodness-of-fit in explanatory models.[40,41] Brier score ranges from 0 to 1, with scores <0.25 indicating better overall performance. Cragg and Uhler's R^2 is a measure of explained model variance and ranges from 0 to 1.[40]

2.3 | Prediction tool

A clinical prediction tool that estimates the probability of a dog receiving a diagnosis of Cushing's syndrome was developed based on the function of the regression coefficients. To derive the points for the predictive tool, the regression coefficients for each predictor variable were used as weights which were divided by a common factor (the smallest significant coefficient in the final model) and rounded to the nearest integer.[42] A dog's total score is calculated by additive combination of the points scored for each predictor.[42,43] The predicted likelihood ($\hat{p}$) for each possible total score was calculated for ease of reference in clinical practice by the following steps:

1. Obtain an estimate of the linear predictor (LP_i) using the rounded points total:

$$LP_i = \beta + B(\text{Points total})$$

(β: optimism-adjusted intercept [constant]; B: common factor).

2. Calculate the predicted likelihood from the inverse logit transformation of the linear predictor:

$$\hat{p} = \frac{e^{LP_i}}{1 + e^{LP_i}}$$

3 | RESULTS

The data set contained 905 544 dogs attending 886 VetCompass participating practices in 2016. Search terms identified 10 141 dogs where Cushing's syndrome was considered as a clinical diagnosis which were manually examined to identify those that fulfilled the criteria for inclusion in the study (n = 1625). Of these, the EHRs of 1000 (61.5%) randomly selected dogs were examined in detail and extraction of clinical information was performed, identifying 419 cases and 581 noncases. Animals with no recorded information regarding clinical signs within the 2-week period of first suspicion were excluded from the study, retaining 398/419 (95.0%) cases and 541/581 (93.1%) noncases for analysis. The final disorders for noncases recorded within the EHR were reported (Table 1). "Endocrine disorders" formed the most common disorder category for noncases (n = 85, 15.7%) and "unspecified hepatic disorder" was the most commonly recorded diagnosis (n = 56, 10.2%). The remaining 8516 dogs were categorized as follows: Cushing's syndrome included as a differential diagnosis term in the EHR but never investigated (n = 4756), confirmed Cushing's cases diagnosed before 2016 (692), suspected iatrogenic Cushing's cases (316), cases with a diagnosis suspected and investigated but never confirmed nor ruled out (1540), an incorrect use of the search terms included (eg, Cushing's suture) (599), or Cushing's syndrome ruled out before 2016 (613).

Median age at first suspicion of Cushing's cases was 10.8 years (IQR 9.0-12.5; range 3.9-17.6) and 10.2 years (IQR 8.2-12.1; range 0.7-18.2) in noncases. Median bodyweight of cases was 11.4 kg (IQR 8.8-20.0; range 2.5-67.0) and 13.2 kg (IQR 9.3-25.1; range 1.7-80.5) in noncases. Of cases, 212 (53.3%) were female compared to 275 (50.5%) noncases. A higher proportion of cases were entire compared to noncases; with 58/398 (14.6%) cases entire females compared to 39/541 (7.2%) noncases and 53 (13.3%) cases were entire males compared to 61 (11.3%) noncases ($P < .01$). Crossbreeds made up 90 cases (22.6%) and 114 noncases (21.1%). The most represented purebred was the Jack Russell terrier (39 cases [9.8%]; 39 noncases [7.2%]), the Staffordshire bull terrier (29 cases [7.3%] and 26 noncases [4.8%]), West Highland white terrier (WHWT) (13 cases [3.3%]; 46 noncases [8.5%]), and the Bichon Frise (32 cases [8.0%]; 24 noncases [4.4%]) (Table 2). Bodyweight was not included in the modeling process as it was considered biologically collinear and therefore inherently related with breed.[28]

Polydipsia was the most commonly recorded clinical sign, present in 540/939 (57.5%) of the study population; 279/398 cases (70.1%) and 261/541 noncases (48.2%) presented with this clinical sign. Polyuria was recorded in 429 (45.7%) of the population, 234/398 (60.2%) of cases and 195/541 (36.0%) of noncases. When comparing dogs presenting with both polydipsia and polyuria, these predictors appeared to be collinear with few cases included in the discordant categories; polydipsia without polyuria was present in 25 dogs (6.3%) and polyuria without polydipsia was present in 136 dogs (25.2%). As the most frequently recorded clinical sign, only polydipsia was included in the modeling process. Vomiting and diarrhea did not appear statistically collinear therefore were both included as potential predictors.

Continuous data for ALKP and USG were not included in analysis as large proportions of the data were missing (>65%) and were not deemed reliable for imputation. When comparing categorized ALKP and ALT predictor variables, >75% of the data was concordant; therefore, these predictors were considered collinear and the most

TABLE 1 Diagnostic terms recorded in the electronic health records for noncase dogs (n = 541) after being suspected of Cushing's syndrome

Disorder category	Noncases (%)	Fine level diagnostic terms (n)
Cardiorespiratory	31 (5.73)	Hypertension (10), bronchitis (7), chronic heart disease (6), pulmonary thromboembolism (3), pericardial effusion (2), cor pulmonale (1), brachycephalic obstructive airway syndrome (1), unspecified respiratory disorder (1)
Dermatological	67 (12.38)	Unspecified dermatological disorder (32), pyoderma (12), alopecia (8), atopy/allergy (6), dermatitis (4), flea allergy dermatitis (3), follicular dysplasia (1), demodicosis (1)
Endocrine	85 (15.71)	Hypothyroidism (34), insulin resistance (24), diabetes mellitus (17), diabetes insipidus (6), diabetic ketoacidosis (1), hyperparathyroidism (2), hypoadrenocorticism (1)
Gastrointestinal	40 (7.39)	Gastroenteritis (34), inflammatory bowel disease (3), parasitic disease (2), megaesophagus (1)
Hepatobiliary	82 (15.16)	Unspecified hepatic disorder (56), hepatitis (12), cholangiohepatitis (11), biliary mucocoele (3)
Infectious/ inflammatory	16 (2.96)	Pancreatitis (12), sepsis (2), peritonitis (1), tooth root abscess (1)
Miscellaneous	42 (7.76)	Transient polydipsia (19), obesity (16), medication adverse effects (4), heat stroke (2), hypertriglyceridemia (1)
Neoplastic	37 (6.84)	Liver mass (9), unspecified mass (8), adrenal mass (5), lymphoma (5), pheochromocytoma (2), brain tumor (2), anal sac carcinoma (1), insulinoma (1), hemangiosarcoma (1), mediastinal mass (1), oral mass (1), transitional cell carcinoma (1)
Neurological	20 (3.70)	Unspecified neurological disorder (11), psychogenic polydipsia (6), cognitive dysfunction (2), idiopathic epilepsy (1)
Ocular	10 (1.85)	Sudden acquired retinal degeneration syndrome (5), nonhealing corneal ulcer (4), keratoconjunctivitis sicca (1)
Orthopedic	22 (4.07)	Arthritis (8), cruciate disease (7), unspecified orthopedic disorder (7)
Renal	25 (4.62)	Chronic kidney disease (18), protein-losing nephropathy (4), proteinuria (3)
Uro-genital	64 (11.83)	Urinary tract infection (28), incontinence (25), urolithiasis (6), prostatic disease (3), unspecified urinary disease (2)

complete variable (ALKP) was included in analysis. Categorized variables of recorded raised ALKP, presence of proteinuria, or low USG were included in model development. Missingness for veterinary reported categorized clinicopathologic data was fairly high at around 50% (Table 2). Clinic ID was not included as a random effect in the final model as the clinic attended accounted for only 1.5% of the variance observed in the data (LRT of rho P = .37, rho = .015).

The final model retained 10 predictors: breed, sex, age, polydipsia, vomiting, potbelly/hepatomegaly, alopecia, pruritus, ALKP, and USG (Table 3). No interactions or additional confounding factors were identified. Age was nonlinearly associated with the outcome and was modeled as linear splines, with cutoffs categorizing age into 3 groups: <7, 7 to <11, and ≥11 years.

Sex and breed were included into the model with entire females and certain breeds (Border terriers and Bichon Frise) associated with an increased predicted likelihood of Cushing's syndrome. Polydipsia, potbelly, and alopecia were associated with an increased predicted likelihood of Cushing's. The presence of a potbelly contributed the greatest increased likelihood of receiving a diagnosis of Cushing's syndrome with an optimism-adjusted coefficient of 0.95 (β-coefficient = 1.11, 95% CI = 0.78-1.43, P < .001). The presence of vomiting and/or pruritus was associated with a reduced predicted likelihood of Cushing's. The presence of a nonelevated ALKP and/or nondilute USG were associated with a reduced predicted likelihood of Cushing's. A nonelevated ALKP had the greatest contribution to reducing the likelihood of receiving a diagnosis of Cushing's syndrome with an optimism-adjusted coefficient of −1.25 (β-coefficient = −1.46, 95% CI = −2.15 to −0.76, P < .001).

3.1 | Model performance

The calibration plot indicated good calibration with the confidence intervals mostly overlapping the 45° line. Higher probability predictions have wider confidence intervals and further deviation from the 45° line of perfect calibration, indicating more uncertainty (Figure 1). The calibration estimates showed a c-slope of 0.86 indicating some overfitting of the model and that predictions were moderately too extreme (ie, low predictions were too low, high predictions were too high) (Table 4). This was corrected for by applying the c-slope value as the shrinkage factor to the model coefficients. CITL of 0.001 indicated that the predictions were systematically well calibrated. Discrimination of the model was relatively good with an AUROC = 0.78 (95% CI = 0.75-0.81) (Figure 2). Optimism-adjusted AUROC was estimated to be 0.76. Brier score was 0.19 and Cragg and Uhler's R^2 was 0.31 indicating moderate overall model performance.

3.2 | Prediction tool

A prediction tool from the final model was developed (Table 5). The smallest significant optimism-adjusted coefficient in the model was used as the common factor to standardize the coefficients and to derive the tool's points, which was for "not recorded USG" (optimism-

TABLE 2 Descriptive statistics and chi-squared associations with gaining a future diagnosis of Cushing's syndrome in dogs attending primary-care veterinary practices in the United Kingdom (cases, n = 398; noncases: n = 541)

Variable	Category	Cases (%)	Noncases (%)	Chi-squared *P* value
Sex	Female entire	58 (14.6)	39 (7.2)	.001
	Female neutered	154 (38.7)	236 (43.6)	
	Male entire	53 (13.3)	61 (11.3)	
	Male neutered	133 (33.4)	205 (37.9)	
Breed	Bichon frise	32 (8.0)	24 (4.4)	<.001
	Border terrier	23 (5.8)	11 (2.0)	
	Crossbreed	90 (22.6)	114 (21.1)	
	Jack Russell terrier	39 (9.8)	39 (7.2)	
	Labrador retriever	6 (1.5)	39 (7.2)	
	Other purebreed	140 (35.2)	198 (36.6)	
	Schnauzer	6 (1.5)	24 (4.4)	
	Staffordshire bull terrier	29 (7.3)	26 (4.8)	
	West highland white terrier	13 (3.3)	46 (8.5)	
	Yorkshire terrier	20 (5.0)	20 (3.7)	
Age (y)	<7	31 (7.8)	93 (17.2)	<.001
	7 to <11	180 (45.2)	229 (42.3)	
	≥11	187 (47.0)	219 (40.5)	
Polydipsia	Yes	279 (70.1)	261 (48.2)	<.001
	No	119 (29.9)	280 (51.8)	
Polyuria	Yes	234 (58.8)	195 (36.0)	<.001
	No	164 (41.2)	346 (64.0)	
Polyphagia	Yes	98 (24.6)	77 (14.2)	<.001
	No	300 (75.4)	464 (85.8)	
Vomiting	Yes	19 (4.8)	59 (10.9)	.001
	No	379 (95.2)	482 (89.1)	
Diarrhea	Yes	26 (6.5)	57 (10.5)	.03
	No	372 (93.5)	484 (89.5)	
Potbelly/hepatomegaly	Yes	197 (49.5)	116 (21.4)	<.001
	No	201 (50.5)	425 (78.6)	
Thin/dry skin	Yes	96 (24.1)	100 (18.5)	.04
	No	302 (75.9)	441 (81.5)	
Alopecia	Yes	118 (29.7)	81 (15.0)	<.001
	No	280 (70.3)	460 (85.0)	
Pruritus	Yes	15 (3.8)	45 (8.3)	.005
	No	383 (96.2)	496 (91.7)	
Muscle wastage	Yes	54 (13.6)	45 (8.32)	.01
	No	344 (86.4)	496 (91.7)	
Lethargy	Yes	73 (18.3)	112 (20.7)	.37
	No	325 (81.7)	429 (79.3)	
Panting	Yes	80 (20.1)	99 (18.3)	.49
	No	318 (79.9)	442 (81.7)	
Neurological signs	Yes	18 (4.5)	31 (5.7)	.41
	No	380 (95.5)	510 (94.3)	
Insulin prescribed	Yes	6 (1.5)	17 (3.1)	.11
	No	392 (98.5)	524 (96.9)	

(Continues)

TABLE 2 (Continued)

Variable	Category	Cases (%)	Noncases (%)	Chi-squared *P* value
Thyroxine prescribed	Yes	14 (3.5)	18 (3.3)	.87
	No	384 (96.5)	523 (96.7)	
Cruciate disease in previous year	Yes	11 (2.8)	7 (1.3)	.10
	No	387 (97.2)	534 (98.7)	
Hospitalized in previous year	Yes	55 (13.8)	81 (15.0)	.62
	No	343 (86.2)	460 (85.0)	
Hypertensive medication prescribed	Yes	3 (0.8)	8 (1.5)	.31
	No	395 (99.2)	533 (98.5)	
Raised ALKP activity	Yes	211 (53.0)	263 (48.6)	.001
	No	14 (3.5)	55 (10.2)	
	Unknown	173 (43.5)	223 (41.2)	
Raised ALT activity	Yes	163 (41.0)	179 (33.1)	<.001
	No	28 (7.0)	98 (18.1)	
	Unknown	207 (52.0)	264 (48.8)	
Low USG	Yes	117 (29.4)	110 (20.3)	.001
	No	49 (12.3)	101 (18.7)	
	Unknown	232 (58.3)	330 (61.0)	
Proteinuria	Yes	95 (23.9)	99 (18.3)	.08
	No	54 (13.6)	90 (16.6)	
	Unknown	249 (62.6)	356 (65.8)	

Abbreviations: ALKP, alkaline phosphatase; ALT, aminotransferase; USG, urine specific gravity.

adjusted coefficient = −0.38). The predicted likelihoods were calculated for each total score to develop a scoring system that covered a range from −13 to 10 (Table 6). An individual dog scoring the lowest possible score of −13 reflects a 0% predicted likelihood and the highest possible score of 10 reflects a 96% predicted likelihood of Cushing's syndrome.

4 | DISCUSSION

Our study outlines the development of a tool that predicts the diagnosis of Cushing's syndrome at the point of first suspicion, using EHRs of dogs under primary-veterinary care in the United Kingdom. The prediction tool has many benefits for veterinarians in primary-care practice. Knowing the predicted likelihood of disease for an individual dog through assimilation of the predictive clinical features of the disease could support decision-making for veterinarians in the practice setting. Using this tool to selectively identify dogs with a higher likelihood of disease before diagnostic testing with a LDDST or ACTH stimulation test could improve the positive predictive value of such tests. For example, using the tool for a 9-year-old, femaled-neutered, crossbreed dog presenting only with polydipsia, ALKP is not elevated and USG is not dilute the dog would gain a score of −3, indicating a 15% predicted likelihood of Cushing's syndrome. In this situation, the attending veterinarian could consider it unlikely the dog has Cushing's at the current time and further testing is not warranted. Additionally, should pituitary-adrenal axis testing have been performed in this case and a positive test obtained, the prediction tool result could highlight that this result carries a low positive predictive value and should not be taken as strong evidence in favor of a diagnosis of Cushing's syndrome. Obtaining a quantitative value of predicted disease likelihood could also aid communication with owners during a consultation and provide transparency of clinical decision-making.

The tool was developed from the prediction model which included clinical signs, demographic factors, and some laboratory factors. The model indicated good discrimination, with an AUROC = 0.78 (95% CI = 0.75-0.81, optimism-adjusted AUROC = 0.76) and a good model fit (Brier score = 0.19 and Cragg-Uhler's R^2 = 0.31). The model largely utilizes the clinical picture and performs well therefore highlighting that gaining a good understanding of the clinical picture is vital.

The predictors assessed in our study were identified a priori based on current knowledge of the disease using existing literature and clinical expertise. The final model retained 10 predictors: breed, sex, age, polydipsia, vomiting, potbelly/hepatomegaly, alopecia, pruritus, ALKP, and USG. The presence of polydipsia and presence of a potbelly contributed a higher predicted likelihood of disease within the models and are commonly associated with Cushing's syndrome in the literature.[4,5,9] Dermatological changes are frequently observed in dogs with Cushing's such as alopecia yet chronic glucocorticoid excess in these dog also means that they are less likely to show signs of pruritus.[6,44] Sex was included in the model with the β-coefficients indicating female-entire dogs had the highest predicted likelihood. The

TABLE 3 Final predictors (including demographic, clinical signs, and clinicopathologic data) for a diagnosis of Cushing's syndrome after multivariable logistic regression with bootstrap resampling, developed in dogs attending primary-care veterinary practices in the United Kingdom (cases, n = 398; noncases: n = 541)

Predictor	Category	β-coefficient	95% Confidence interval	*P* value (Wald)	Optimism-adjusted β-coefficient
Neuter status	Female-entire	Baseline	—	—	—
	Female-neutered	−.64	−1.17 to −0.12	.02	−.55
	Male-entire	−.34	−0.97 to 0.29	.29	−.29
	Male-neutered	−.60	−1.13 to −0.07	.03	−.52
Age (y)	<7	Baseline	—		—
	7 to <11	.64	0.13-1.15	.01	.55
	≥11	.58	0.06-1.09	.03	.50
Polydipsia	Yes	.87	0.55-1.20	<.001	.75
	No	Baseline	—		—
Vomiting	Yes	−.76	−1.37 to −0.14	.02	−.65
	No	Baseline	—		—
Potbelly	Yes	1.11	0.78-1.43	<.001	.95
	No	Baseline	—		—
Alopecia	Yes	.94	0.54-1.33	<.001	.80
	No	Baseline	—		—
Pruritus	Yes	−0.88	−1.56 to −0.20	.01	−.76
	No	Baseline	—		—
Breed	Crossbreed	Baseline	—		—
	Bichon frise	.68	0.01 to 1.35	.05	.58
	Border terrier	.61	−0.26 to 1.48	.17	.52
	Jack Russell terrier	.11	−0.47 to 0.69	.72	.09
	Labrador retriever	−1.37	−2.33 to −0.42	.005	−1.18
	Other purebred	−.04	−0.44 to 0.36	.84	−.04
	Schnauzer	−1.03	−2.06 to 0.01	.05	−.88
	Staffordshire bull terrier	.05	−0.63 to 0.73	.47	.04
	West Highland white terrier	−1.18	−1.91 to −0.45	.001	−1.02
	Yorkshire terrier	.09	−0.70 to 0.88	.82	.08
USG	Dilute	Baseline	—	—	—
	Not dilute	−.85	−1.35 to −0.36	.001	−.73
	Not recorded	−.43	−0.82 to −0.06	.02	−.38
ALKP	Elevated	Baseline	—	—	—
	Not elevated	−1.46	−2.15 to −0.76	<.001	−1.25
	Not recorded	−.16	−0.48 to 0.17	.34	−.13
Constant		−.49	−1.24 to 0.25	.19	−.42

Note: β-coefficients were multiplied by the optimism-adjusted calibration-slope (0.86), estimated through bootstrap resampling to produce optimism-adjusted coefficients.
Abbreviations: ALKP, alkaline phosphatase; USG, urine specific gravity.

reason for this observation is not known. A sex predisposition for Cushing's syndrome has been investigated in studies examining this causal relationship, with no clear association determined.[2,16,45] However it must be reiterated that the primary aim of the current study was to describe the predictive rather than the causal relationships between the 2 groups being investigated.[46] Our study includes different comparative populations of dogs to previous studies that specifically looked for causal relationships. Breeds such as the WHWT and Labrador retriever had low predicted likelihood of Cushing's syndrome. These breeds could have been overrepresented in the noncases because of predisposition for other diseases presenting in a similar way to Cushing's. For example, WHWTs are predisposed to skin disease and might have had Cushing's investigated because of a dermatology work up.[47]

Raised ALKP has been frequently reported in dogs with a diagnosis of Cushing's syndrome.[9,48] Additionally, a low USG has often been recorded in the literature.[49,50] ALKP and USG were included as

categorical variables because of poor recording of specific values in the EHRs. Availability of clinicopathologic data in our study was reliant on the test having been performed within primary-care practice and dependent on the system used by the veterinary practice to record this information. Some tests performed at external laboratories were not captured within VetCompass limiting the inclusion of this data into the study. Additionally, variations in the laboratory equipment used across practices might have introduced some noise into the analysis of these factors. Further refinement to include additional clinicopathologic data could provide future direction for this tool, such as cholesterol and the stress leukogram, which were infrequently reported in the EHRs and therefore not considered during data extraction in our study. Multiple imputation has been shown to be unbiased for estimating missing data up to 50% but can become unreliable for certain types of missingness such as if data are "missing not at random" and associated with the outcome of interest.[30,31] It was elected not to impute this data and instead these data were included within a "not recorded" category for the ALKP and USG predictor variables rather than excluding these variables or performing a complete case analysis.

Inappropriate prediction model development can lead to poor model fit, giving falsely high and "optimistic" results which do not perform well in novel data sets.[18,19,51,52] Therefore, a necessary part of model development is internal validation.[18] Resampling techniques such as bootstrapping are recommended, as opposed to split sampling methods, as they optimize data usage to enable the model to be developed and internally validated on the whole data set without losing any predictive power.[33,53] Bootstrap resampling estimates of performance indicate how the results will generalize to an independent data set derived from the same population.[33] The optimism-adjusted estimates, which account for potential overfitting and are less "optimistic," showed good performance. There was some overfitting of the model, indicated by the calibration slope and the calibration plot indicated weaker calibration at the higher probability predictions. The overfitting was accounted for by shrinking the model coefficients therefore these adjusted estimates

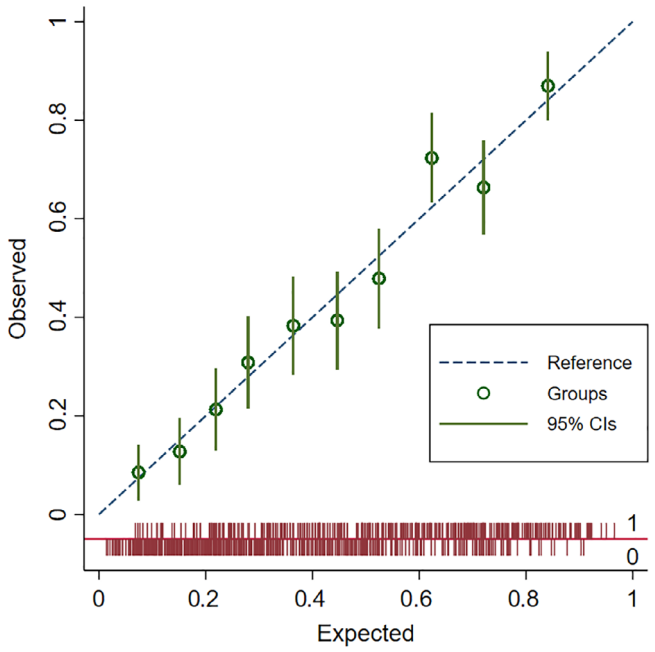

FIGURE 1 Calibration plot of the final prediction model for a diagnosis of Cushing's syndrome using multivariable logistic regression with bootstrap resampling, developed in dogs attending primary-care veterinary practices in the United Kingdom (cases, n = 398; noncases: n = 541). The plot describes the mean observed proportions of dogs with a diagnosis of Cushing's compared to the mean predicted probabilities, by deciles of predictions. The 45° line denotes perfect calibration

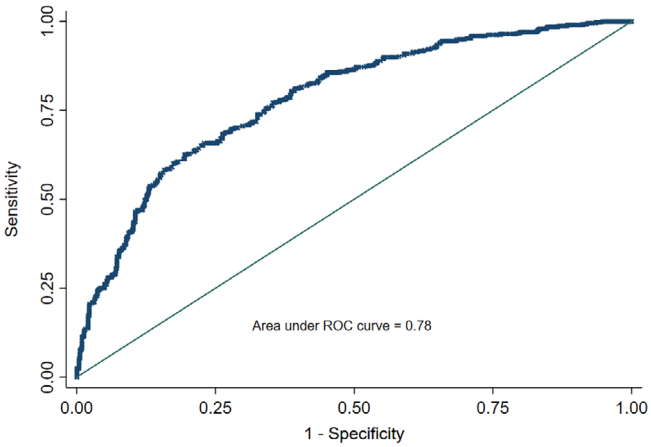

FIGURE 2 Receiver operating characteristic curve for the final prediction model for a diagnosis of Cushing's syndrome using multivariable logistic regression with bootstrap resampling, developed in dogs attending primary-care veterinary practices in the United Kingdom (cases, n = 398; noncases: n = 541)

	Apparent performance	Average optimism	Optimism-adjusted performance
AUROC	0.78	0.02	0.76
CITL	0.00	−0.001	0.001
C-slope	1.00	0.14	0.86

Abbreviations: AUROC, area under the receiver operating characteristic curve; CITL, calibration-in-the-large.

TABLE 4 Apparent performance measures (performance of the bootstrap samples), average optimism (the average difference between the performance of the bootstrap samples and the dogs not included in the bootstrap samples) and estimated optimism-adjusted performance measures for the final model (cases, n = 398; noncases: n = 541)

TABLE 5 Prediction tool to calculate the likelihood of a dog having Cushing's syndrome using demographic, clinical sign, and laboratory predictive factors, developed in dogs attending primary-care veterinary practices in the United Kingdom (cases, n = 398; noncases: n = 541)

	Category	Points	Points scored
Dog demography			
Neuter status	Female-entire	0	
	Female-neutered	−1	
	Male-entire	−1	
	Male-neutered	−1	
Current age (years)	<7	0	
	≥7	1	
Breed	Bichon frise	2	
	Border terrier	1	
	Labrador retriever	−3	
	Schnauzer	−2	
	West Highland white terrier	−3	
	Other breed or crossbreed	0	
Presenting clinical signs			
Polydipsia	Yes	2	
	No	0	
Vomiting	Yes	−2	
	No	0	
Potbelly/hepatomegaly	Yes	3	
	No	0	
Alopecia	Yes	2	
	No	0	
Pruritus	Yes	−2	
	No	0	
Laboratory factors			
Urine specific gravity	Dilute (≤ 1.020)	0	
	Not dilute (> 1.020)	−2	
	Not recorded	−1	
Serum ALKP	Elevated	0	
	Not elevated	−3	
	Not recorded	0	
		Total score:	

Note: Regression β-coefficients from model B for each predictor variable were used as weights which were multiplied by a common factor ("Not recorded" USG optimism-adjusted coefficient = 0.38) and rounded to the nearest integer. To calculate the predicted likelihood of an individual dog having Cushing's syndrome, add together the points that correspond to the category for each predictor and match to the Table 6 below.

Abbreviations: ALKP, alkaline phosphatase; USG, urine specific gravity.

are likely representative of the tools performance in primary-care practice. For models to be clinically useful, it is vital they are developed in a large, representative sample of the target population of interest to optimize their predictive performance.[17,18] The dogs included in our study were selected from the largest research database of primary-care EHRs in the United Kingdom, representing approximately 30% of all UK veterinary practices.[22] However, the tool could be applied to different prevalence populations when used in practice and could have some impact on the predictive performance demonstrated in our study. Future external validation (using an external, independent data set) of the final prediction tool is required to assess its wider generalizability and performance in clinical practice.[17,39,54]

The predictor variables included in our study represent the clinical information typically used by veterinarians in practice to formulate a perceived "pretest" probability of disease.[55,56] The dogs included in our study were required to have been suspected of having Cushing's in the EHRs therefore were presumably perceived to have a greater "pretest" probability of Cushing's by the attending veterinarian. The tool could perform differently at varying thresholds of "pretest" probabilities, used by differing veterinarians to consider the animal as a

TABLE 6 Points total and predicted likelihood of an individual dog having Cushing's syndrome using demographic, clinical sign, and laboratory predictive factors, developed in dogs attending primary-care veterinary practices in the United Kingdom (cases, n = 398; noncases: n = 541)

Points total	Predicted likelihood of Cushing's syndrome (0.00 = 0%; 0.96 = 96%)
−13	0.00
−12	0.01
−11	0.01
−10	0.01
−9	0.02
−8	0.03
−7	0.04
−6	0.05
−5	0.08
−4	0.11
−3	0.15
−2	0.20
−1	0.27
0	0.35
1	0.44
2	0.53
3	0.63
4	0.71
5	0.78
6	0.84
7	0.88
8	0.92
9	0.94
10	0.96

Note: The linear predictor (LP_i) using the rounded points total was estimated: $LP_i = \beta + B(\text{Points total})$ (β: optimism-adjusted intercept [constant]; *B*: common factor). Then the predicted likelihood from the inverse logit transformation of the linear predictor was calculated: $\hat{p} = \frac{e^{LP_i}}{1+e^{LP_i}}$.

potential Cushing's case. Clinical decision-making includes lots of uncertainty as the perceived "pretest" probability formulated by the veterinarian is subjective and likely varies between clinicians.[55] The tool developed in our study aimed to reduce some uncertainty surrounding this clinical decision by helping to standardize the diagnostic approach, without removing the clinical freedom of decision-making by the veterinarian.

Adequate sample size is also important when developing a prediction tool, with small samples leading to spurious associations from overfitting of the data, producing coefficients that are too large and a model that is too extreme.[24,25] An a priori sample size estimation was carried out to increase confidence that an adequate proportion of cases were manually reviewed for inclusion in the study. Sample size estimation using the EPV criteria of 10 cases per variable is frequently cited in the literature; however, the reliability of this to ensure adequate sample size has been questioned and other more reliable methods are warranted.[57]

The use of strict inclusion criteria was used to increase confidence and minimize misclassification between the identification of the cases of Cushing's syndrome and the control population. There will have been some dogs with Cushing's syndrome in the underlying denominator population that did not meet the study definitions and were excluded from analysis, highlighting the realities of primary-care practice and the importance of using the intended population in the development of a disease prediction tool.[58] Additionally the categorization of dogs as either a case or noncase were based on the diagnosis recorded within the EHR by the attending veterinarian; therefore, this could have introduced some misclassification bias.

The developed tool makes a prediction of diagnosis early in the trajectory of the disease. This was done to assist the diagnostic process at the time point when veterinarians are making this clinical decision in practice and this was standardized for cases and noncases, reducing the potential for bias in the identification of candidate predictors. Clinical signs present within a 2-week window from the point of first suspicion were recorded to keep the dogs' clinical presentation precise to that particular time frame and to reduce influence of bias from the clinician with increasing or decreasing suspicion as disease investigation progresses. When recording clinical sign data from the EHRs, an assumption was made that if the clinical sign was present within the 2-week window, it was likely to have been reported.[59] It was deemed unlikely that veterinarians would routinely record absent clinical signs and therefore omission of a clinical sign was recorded as "not present." The requirement to have at least 1 clinical sign recorded in the EHRs at the time of first suspicion was included to remove cases that might have been poorly recorded. If at least 1 clinical sign was recorded, it is assumed that this was the clinical sign of most concern to the owner and/or the vet and therefore contributed to the decision to undertake further investigations. In our study, polydipsia was recorded in 70% of cases and polyuria was recorded in 59%. With these 2 clinical signs inherently related and few dogs recorded with discordant clinical signs, this could suggest that certain clinical signs are more frequently and accurately recorded in the EHRs at primary-care practices. Therefore, it is possible that some clinical signs were present but remained unnoticed by the owner and not explicitly recorded by the veterinarian during the consultation. These assumptions could have resulted in some misclassification of the clinical signs status; however, any misclassification is likely similar for cases and noncases so could bias the results to the null.

There are some limitations to our study. Some survival bias could have been introduced by including incident cases between January 1, 2016 and June 1, 2018, with some dogs not surviving for the entire study period and reducing their chance of being included in the study. This potential bias is likely small over 2 years and likely similar for cases and noncases. This study period was chosen to avoid excluding dogs with a longer period of disease investigation and to reduce the number of dogs where a diagnosis of Cushing's is neither confirmed nor ruled out. Enhanced methods for case finding and selection would be beneficial to extract greater volumes of information from such

large databases of EHRs. Novel, computationally intensive methods such as natural language processing are being developed to facilitate the identification of larger numbers of cases from the broader denominator population.[60] Additionally as data were retrospective and not recorded primarily for research purposes, there could be variations in how information was recorded by different veterinarians which could introduce noise and reduce the performance of the score.

In conclusion, our study demonstrates the development of a diagnostic prediction tool for Cushing's syndrome in dogs at the point of first suspicion. The tool provided takes a dog's demography, presenting clinical signs and some routine laboratory results into consideration and demonstrated a good predictive performance. The tool can immediately be utilized in primary-care practice to directly aid clinical decision-making and increase confidence in diagnosis. Development of similar tools could prove beneficial for similarly hard to diagnose conditions and it is hoped that this will ultimately result in a positive impact on animal welfare.

ACKNOWLEDGMENTS

We are grateful to Dechra Veterinary Products Ltd for their funding of this research. We acknowledge the Medivet Veterinary Partnership, Vets4Pets/Companion Care, Goddard Veterinary Group, Independent Vet Care, CVS Group, Beaumont Sainsbury Animal Hospital, Vets Now, and the other UK practices who collaborate in VetCompass. We are grateful to The Kennel Club and The Kennel Club Charitable Trust for supporting VetCompass.

CONFLICT OF INTEREST DECLARATION

I. S. is supported at the RVC by an award from Dechra Veterinary Products Ltd. S. J. M. N. has undertaken consultancy work for Dechra Veterinary Products Ltd. The remaining authors have no conflicts of interest to declare.

OFF-LABEL ANTIMICROBIAL DECLARATION

Authors declare no off-label use of antimicrobials.

INSTITUTIONAL ANIMAL CARE AND USE COMMITTEE (IACUC) OR OTHER APPROVAL DECLARATION

Approval granted by the RVC Ethics and Welfare Committee (URN SR2018-1652).

HUMAN ETHICS APPROVAL DECLARATION

Authors declare human ethics approval was not needed for this study.

ORCID

Imogen Schofield https://orcid.org/0000-0003-3169-8723
Dan G. O'Neill https://orcid.org/0000-0003-1115-2723

REFERENCES

1. O'Neill DG, Scudder C, Faire JM, et al. Epidemiology of hyperadrenocorticism among 210,824 dogs attending primary-care veterinary practices in the UK from 2009 to 2014. *J Small Anim Pract*. 2016;57:365-373.
2. Carotenuto G, Malerba E, Dolfini C, et al. Cushing's syndrome—an epidemiological study based on a canine population of 21,281 dogs. *Open Vet J*. 2019;9:27-32.
3. Helm JR, McLauchlan G, Boden LA, et al. A comparison of factors that influence survival in dogs with adrenal-dependent hyperadrenocorticism treated with mitotane or trilostane. *J Vet Intern Med*. 2011;25:251-260.
4. Fracassi F, Corradini S, Floriano D, et al. Prognostic factors for survival in dogs with pituitary-dependent hypercortisolism treated with trilostane. *Vet Rec*. 2015;176:49.
5. Nagata N, Kojima K, Yuki M. Comparison of survival times for dogs with pituitary-dependent hyperadrenocorticism in a primary-care hospital: treated with trilostane versus untreated. *J Vet Intern Med*. 2017;31:22-28.
6. Behrend EN, Kooistra HS, Nelson R, Reusch CE, Scott-Moncrieff JC. Diagnosis of spontaneous canine hyperadrenocorticism: 2012 ACVIM consensus statement (small animal). *J Vet Intern Med*. 2013;27:1292-1304.
7. Nivy R, Refsal KR, Ariel E, Kuzi S, Yas-Natan E, Mazaki-Tovi M. The interpretive contribution of the baseline serum cortisol concentration of the ACTH stimulation test in the diagnosis of pituitary dependent hyperadrenocorticism in dogs. *J Vet Intern Med*. 2018;32:1897-1902.
8. Monroe WE, Panciera DL, Zimmerman KL. Concentrations of noncortisol adrenal steroids in response to ACTH in dogs with adrenal-dependent hyperadrenocorticism, pituitary-dependent hyperadrenocorticism, and nonadrenal illness. *J Vet Intern Med*. 2012;26:945-952.
9. Bennaim M, Shiel RE, Forde C, Mooney CT. Evaluation of individual low-dose dexamethasone suppression test patterns in naturally occurring hyperadrenocorticism in dogs. *J Vet Intern Med*. 2018;32:967-977.
10. Kaplan AJ, Peterson ME, Kemppainen RJ. Effects of disease on the results of diagnostic tests for use in detecting hyperadrenocorticism in dogs. *J Am Vet Med Assoc*. 1995;207:445-451.
11. Van Liew CH, Greco DS, Salman MD. Comparison of results of adrenocorticotropic hormone stimulation and low-dose dexamethasone suppression tests with necropsy findings in dogs: 81 cases (1985-1995). *J Am Vet Med Assoc*. 1997;211:322-325.
12. Schofield I, Brodbelt DC, Wilson ARL, Niessen S, Church D, O'Neill D. Survival analysis of 219 dogs with hyperadrenocorticism attending primary care practice in England. *Vet Rec*. 2020;186:348-357.
13. Rijnberk A, Mol J. Assessment of two tests for the diagnosis of canine hyperadrenocorticism. *Vet Rec*. 1988;122:178-180.
14. Smiley LE, Peterson ME. Evaluation of a urine cortisol:creatinine ratio as a screening test for hyperadrenocorticism in dogs. *J Vet Intern Med*. 1993;7:163-168.
15. Schofield I, O'Neill DG, Brodbelt DC, et al. Development and evaluation of a health-related quality-of-life tool for dogs with Cushing's syndrome. *J Vet Intern Med*. 2019;33:2595-2604.
16. Hoffman JM, Lourenço BN, Promislow DEL, Creevy KE. Canine hyperadrenocorticism associations with signalment, selected comorbidities and mortality within North American veterinary teaching hospitals. *J Small Anim Pract*. 2018;59:681-690.
17. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:1-73.
18. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med*. 2015;13:1-54.
19. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*. 2008;117:743-753.
20. Jammeh EA, Carroll CB, Pearson SW, et al. Machine-learning based identification of undiagnosed dementia in primary care: a feasibility study. *BJGP Open*. 2018;2:4-12.
21. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9:103.

22. VetCompass [Internet]. Veterinary Companion Animal Surveillance System; c2020. https://www.rvc.ac.uk/vetcompass. Accessed April 17, 2020.
23. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res*. 2017;26:796-808.
24. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373-1379.
25. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, et al. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making*. 2001;21:45-56.
26. Acierno MJ, Brown S, Coleman AE, et al. ACVIM consensus statement: guidelines for the identification, evaluation, and management of systemic hypertension in dogs and cats. *J Vet Intern Med*. 2018;32:1803-1822.
27. Rewerts J, Grooters A, Payne J, Kornegay JN. Atraumatic rupture of the gastrocnemius muscle after corticosteroid administration in a dog. *J Am Vet Med Assoc*. 1997;210:655-657.
28. Dohoo IR, Martin W, Stryhn H. *Veterinary Epidemiologic Research*. 2nd ed. Charlottetown, Canada: VER Inc; 2012.
29. Katz MH. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers*. Cambridge, UK: Cambridge University Press; 2011.
30. Lee KJ, Carlin JB. Recovery of information from multiple imputation: a simulation study. *Emerg Themes Epidemiol*. 2012;9:3-12.
31. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581-592.
32. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Boca Raton,, FL: CRC Press; 1994.
33. Steyerberg EW, Harrell FE Jr, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
34. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26:5512-5528.
35. Sun G-W, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. 1996;49:907-916.
36. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361-387.
37. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. New York, NY: Springer; 2015.
38. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Anaylsis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester, UK: John Wiley & Sons; 2008.
39. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35:1925-1931.
40. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21:128-142.
41. Fenlon C, O'Grady L, Doherty ML, Dunnion J. A discussion of calibration techniques for evaluating binary and categorical predictive models. *Prev Vet Med*. 2018;149:107-114.
42. Mehta HB, Mehta V, Girman CJ, Adhikari D, Johnson ML. Regression coefficient-based scoring system should be used to assign weights to the risk index. *J Clin Epidemiol*. 2016;79:22-28.
43. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med*. 2004;23:1631-1660.
44. Frank LA. Comparative dermatology—canine endocrine dermatoses. *Clin Dermatol*. 2006;24:317-325.
45. Gallelli MF, Cabrera Blatter MF, Castillo V. A comparative study by age and gender of the pituitary adenoma and ACTH and alpha-MSH secretion in dogs with pituitary-dependent hyperadrenocorticism. *Res Vet Sci*. 2010;88:33-40.
46. Shmueli G. To explain or to predict? *Stat Sci*. 2010;25:289-310.
47. O'Neill DG, Ballantyne ZF, Hendricks A, Church DB, Brodbelt DC, Pegram C. West Highland White Terriers under primary veterinary care in the UK in 2016: demography, mortality and disorders. *Canine Genet Epidemiol*. 2019;6:7-18.
48. Mawby DI, Whittemore JC, Fecteau KA. Canine pancreatic-specific lipase concentrations in clinically healthy dogs and dogs with naturally occurring hyperadrenocorticism. *J Vet Intern Med*. 2014;28:1244-1250.
49. Neiger R, Ramsey I, O'Connor J, et al. Trilostane treatment of 78 dogs with pituitary-dependent hyperadrenocorticism. *Vet Rec*. 2002;150:799-804.
50. Ruckstuhl NS, Nett CS, Reusch CE. Results of clinical examinations, laboratory tests, and ultrasonography in dogs with pituitary-dependent hyperadrenocorticism treated with trilostane. *Am J Vet Res*. 2002;63:506-512.
51. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ*. 2007;335:136.
52. Perel P, Arango M, Clayton T, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ*. 2008;336:425-429.
53. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *J R Stat Soc*. 1999;48:313-329.
54. Debray TP, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68:279-289.
55. Cipoli DE, Martinez EZ, Castro M, et al. Clinical judgment to estimate pretest probability in the diagnosis of Cushing's syndrome under a Bayesian perspective. *Arq Bras Endocrinol Metabol*. 2012;56:633-637.
56. Bianchi MT, Alexander BM, Cash SS. Incorporating uncertainty into medical decision making: an approach to unexpected test results. *Med Decis Making*. 2009;29:116-124.
57. van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: beyond events per variable criteria. *Stat Methods Med Res*. 2019;28:2455-2474.
58. Bradley R, Tagkopoulos I, Kim M, et al. Predicting early risk of chronic kidney disease in cats using routine clinical laboratory tests and machine learning. *J Vet Intern Med*. 2019;33:2644-2656.
59. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2017;24:198-208.
60. Kennedy N, Brodbelt DC, Church DB, O'Neill DG. Detecting false-positive disease references in veterinary clinical notes without manual annotations. *npj Digit Med*. 2019;2:33-42.

How to cite this article: Schofield I, Brodbelt DC, Niessen SJM, et al. Development and internal validation of a prediction tool to aid the diagnosis of Cushing's syndrome in dogs attending primary-care practice. *J Vet Intern Med*. 2020; 34:2306–2318. https://doi.org/10.1111/jvim.15851